DOCUMENT RESUME

ED 129 847 TM 005 467

AUTHOR Koffler, Stephen L.
TITLE An Evaluation and Comparison of Discrimination Procedures for Certain Types of Non-Normal Distributions.
PUB DATE [Apr 76]
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Behavioral Science Research; Classification; *Comparative Analysis; *Discriminant Analysis; Hypothesis Testing; *Nonparametric Statistics; Observation; Probability; *Statistical Analysis
IDENTIFIERS Density Estimators; Linear Discriminant Function; Monte Carlo Methods; Nearest Neighbor with Probability Blocks; Quadratic Discriminant Function

ABSTRACT

The power of the classical Linear Discriminant Function (LDF) is compared, using Monte Carlo techniques with five other procedures for classifying observations from certain non-normal distributions. The alternative procedures considered are the Quadratic Discriminant Function, a Nearest Neighbor Procedure with Probability Blocks, and three density estimators. Comparisons of misclassifications are examined for varying sample sizes for two and three dimensional models. Three types of distributions are considered: finite range (Logit Normal), semi-infinite range (Log Normal), and infinite range (Inverse Hyperbolic Sine Normal). Results indicate that certain alternatives to the LDF classify observations correctly in a greater proportion than does the LDF for non-normal data, and that different procedures are best for different types of non-normality. (Author)

AN EVALUATION AND COMPARISON OF DISCRIMINATION PROCEDURES FOR CERTAIN TYPES OF NON-NORMAL DISTRIBUTIONS

Stephen L. Koffler
New Jersey Department of Education

Paper presented at the 1976 annual meeting
of the American Educational Research Association

Introduction

The problem of discrimination (or classification) has always been one of major concern to the behavioral scientist, and one for which there has not always been a satisfactory solution. The discrimination problem arises when the researcher must rationally assign an individual or object to one of a finite number of populations on the basis of a series of measurements obtained on that individual or object, as well as any other pertinent information available.

Fisher (1936) presented the first clear solution to the classification problem. Fisher's solution, called the Linear Discriminant Function (LDF), was the linear combination of the measurements which maximized the ratio of the difference between the sample means to the standard deviation within samples. The inception of a theoretical solution to the problem emerged when the hypothesis testing concepts of Neyman and Pearson were adapted to the discrimination problem by Welch (1939). Welch noted that discrimination procedures classifying new p-dimensional observations were equivalent to partitionings of the sample space into mutually exclusive and exhaustive regions R_i.

For a specific partitioning of the sample space, the observation Z to be classified is assigned to population S_i if the p-dimensional point lies in region R_i. Using this rationale, there are

two possible types of errors that can be committed for the two population problem when classifying the observations.

1. The observation can be classified as originating from population S_1 when it actually comes from S_2.
2. The observation can be classified as originating from population S_2 when it actually comes from S_1.

Associated with each of these errors is a probability of committing the error (called the probabilities of misclassification). Welch (1939) proposed that the optimum classification procedure be that procedure which partitions the sample space in such a manner that the corresponding probabilities of misclassification are minimized.

Welch showed that the optimal partitioning is effected by forming the ratio of the densities of the two populations, $f_1(x)/f_2(x)$. The observation to be classified is assigned to population S_1 if the value of the likelihood ratio is greater than some appropriately determined constant k, and the observation is assigned to population S_2 if the value of the likelihood ratio is less than k.

Anderson (1951) has shown that the discrimination problem can be thought of as a problem of "statistical decision functions": There are a finite number of hypotheses, each hypothesis stating that the distribution of the observation is a specified one; one of the hypothesis is not rejected, the remainder are. Anderson (1951) has shown that a good classification procedure is one which minimizes the "cost"(loss function) of misclassification associated with the procedure. The decision theoretic objective is one of determining an appropriate classification rule that will minimize the risk (expected loss) associated with the procedure.

The procedure that minimizes the risk function for given a priori

probabilities (a priori probability q_i that the observation to be classified belongs to population S_i) is a Bayes procedure. When the a priori probabilities are not known, Von Mises (1945) has shown that the procedure which allows the maximum of the minimum probability of correct classification (the minimax procedure) effects the best partitioning of the sample space.

In either situation (whether the a priori probabilities are known or unknown), the form of the best classification procedure is the ratio of the density functions of the populations. When the a priori probabilities are equal and the costs of misclassification are equal, the best procedure is such that an observation Z is classified as belonging to S_1 if $f_1(z)/f_2(z) > 1$ and Z is classified into S_2 if $f_1(z)/f_2(z) < 1$. When the likelihood ratio is identical to one, a randomized procedure is used to classify the observations.

Frequently behavioral scientists collect data that is representative of multivariate normal distributions. For the two population situation, when the populations are multivariate normal with a common variance-covariance matrix and known mean vectors, the optimal classification procedure (the ratio of the densities) simplifies to:

$$L = \underset{\sim}{Z}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) - 1/2(\underset{\sim}{\mu}_1+\underset{\sim}{\mu}_2)'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) \qquad (1)$$

The first term of (1) is that to which Fisher's LDF reduces if the population means and common variance-covariance matrix are known. Anderson (1951) has shown that the optimal probability of misclassification associated with (1) assuming equal costs of misclassification and equal a priori probabilities, is $\Phi(-\Delta/2)$, where $\Delta^2 = (\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$ and $\Phi(\cdot)$ is the ordinate of the normal distribution function.

When the variance-covariance matrices of the two multivariate normal populations are not identical, the form of the likelihood ratio is not a linear function, but instead a quadratic function, called the Quadratic Discriminant Function.

$$QDF = 1/2\underset{\sim}{z}'(\underset{\sim}{\Sigma}_1^{-1}-\underset{\sim}{\Sigma}_2^{-1})\underset{\sim}{z} + (\underset{\sim}{\mu}_2'\underset{\sim}{\Sigma}_2^{-1} - \underset{\sim}{\mu}_1'\underset{\sim}{\Sigma}_1^{-1})\underset{\sim}{z} + 1/2([\underset{\sim}{\mu}_1'\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2'\underset{\sim}{\Sigma}_2^{-1}\underset{\sim}{\mu}_2] - \log[\underset{\sim}{\Sigma}_2/\underset{\sim}{\Sigma}_1]) \quad (2)$$

It is not unusual that researchers must consider situations in which the distributions from which their data are drawn are not completely specified, but instead are known except for one or more parameters. When these circumstances arise, the unknown parameters of the distributions must be estimated from samples. Then classification procedures are developed which are based on the sample estimates.

To determine appropriate sample based classification procedures it is appropriate to select those procedures whose risk functions asymptotically approach the risk function of the optimal procedure, (i.e., those sample based procedures which are consistent). Hoel and Peterson (1949) intuitively reasoned that the best sample based procedure would be of the type of the likelihood ratio in which the sample estimates replaced the unknown parameters (called a "plug-in" procedure). Fix and Hodges (1951) showed that the "plug-in" procedure of Hoel and Peterson was a consistent procedure and was the most appropriate sample based technique to use.

For the case in which the parameters of the multivariate normal distribution are unknown, Anderson (1951) developed a statistic (W) which is the "plug-in" analogue to the LDF. The W statistic is of

the form of (1) in which the maximum likelihood estimates replace the unknown parameters. The first term of Anderson's W statistic is the form of the LDF first obtained by Fisher (1936). Using similar arguments, the form of the QDF when based on sample estimates is identical to (2) with $\bar{\underset{\sim}{X}}_i$ replacing $\underset{\sim}{\mu}_i$ and $\underset{\sim}{S}_i$ replacing $\underset{\sim}{\Sigma}_i$.

Historically, the Linear Discriminant Function (or the Anderson W statistic) has been used almost exclusively for discrimination problems regardless of whether the assumption of multivariate normality of the underlying populations has been satisfied. However, Lachenbruch, Sneering, and Revo (1973) have shown that the LDF is clearly not a robust procedure when used to classify observations from non-normal distributions. Because of this it is inappropriate to use the LDF with data that is not representative of a multivariate normal distribution. When the samples are drawn from populations that are of some known distribution, the optimal procedure is the ratio of the densities. In most situations, however, information is obtained from samples drawn from unknown populations and alternative distribution-free classification methods should be employed.

The most desirable type of nonparametric classification procedure is a procedure which is consistent with the likelihood ratio procedure. Fix and Hodges (1951) considered a solution to the nonparametric discrimination problem based on estimates of the unknown densities, and used these estimates as "plug-in" versions of the likelihood ratio procedure. Alternative types of nonparametric classification procedures suggested include Nearest Neighbor type procedures and certain methods based nonparametric rank tests.

It was the purpose of the research described in this paper to

empirically contrast the discriminatory power of alternative two population classification procedures to the classical LDF when classifying data that originates from certain types of non-normal distributions.

Model and Methodology

Let X_1 and X_2 be two absolutely continuous p-dimensional random variables; their probability density function given by $f_1(x)$ and $f_2(x)$, respectively. Using Monte Carlo techniques, samples from four types of p-dimensional distributions, all of whose dimensions were independent, were generated for the two population discrimination problem. The four distributions included the multivariate normal distribution and non-normal representatives from three classes of distribution: 1) distributions with finite range; 2) distributions with semi-infinite range; and 3) distributions with infinite range.

The three non-normal distributions were generated from the Johnson (1949) system of distributions. The distributions were the Log Normal distribution, the Logit Normal distribution, and the Inverse Hyperbolic Sine Normal distribution. To obtain the required non-normal samples, normally distributed random variables were generated and the appropriate inverse transformation performed. The Johnson system of transformations and inverse transformations is summarized in Table 1.

In Table 1, the variable y is normally distributed with given mean and variance; the variable x is distributed according to the appropriate non-normal distribution. To obtain random points from a normal distribution, uniform random deviates were generated from

the IBM Scientific Subroutine Package. Then using the Central Limit Theorem, the normally distributed random data points were determined.[1]

TABLE 1

TRANSFORMATIONS (AND THEIR INVERSES)
THAT GENERATE THE JOHNSON (1949)
SYSTEM OF DISTRIBUTIONS

Distribution	Transformation		Inverse
Log Normal	$y=\log x$	$0\leq x\leq\infty$	$x=EXP(y)$
Logit Normal	$y=\log(x/1-x)$	$0<x<1$	$x=EXP(y/1-y)$
Inverse Hyperbolic Sine Normal	$y=\mathrm{Sinh}^{-1}(x)$	$-\infty<x<\infty$	$x=\mathrm{Sinh}(y)$

The p-dimensional normal distributions which were used to generate the non-normal distributions had the identity matrix as their variance-covariance matrix. The mean vector for population S_1 was $(\mu, 0, \ldots, 0)$ and for population S_2, $(0, \ldots, 0)$. For each of the four distributions, samples were generated for each combination of sample size ($n = 64, 200, 729$), first component of the mean vector for population S_1 ($\mu = 1,2,3$), and dimensionality ($p = 2,3$).

For each of the eighteen possible combinations of the parameters for each of the four distributions, six different classification rules were developed. The classification procedures considered were:

1. Linear Discriminant Function (Anderson W Statistic)
2. Quadratic Discriminant Function

[1]There has been in the past some criticism leveled as to the validity of IBM's Scientific Subroutine Package random number generator. Using Chi-Square tests described by MacLaren and Marsaglia (1965), it was determined that the uniform random deviates generated from the program were random. Further, using the Kolmogorov-Smirnov test, it was determined that the random variables obtained from the uniform random deviates were representatives of a normal distribution.

3. Nearest Neighbor with Probability Blocks
4. Parzen-Cacoullos Density Estimator
5. Loftsgaarden-Quesenberry Density Estimator
6. Gessaman Density Estimator

Linear Discriminant Function

The Anderson W Statistic was used, assuming equal a priori probabilities and equal costs of misclassification. For this situation, the Anderson W statistic is of the form:

$$W = \underset{\sim}{Z}'\underset{\sim}{S}^{-1}(\bar{\underset{\sim}{X}}_1-\bar{\underset{\sim}{X}}_2) - 1/2(\bar{\underset{\sim}{X}}_1+\bar{\underset{\sim}{X}}_2)'S^{-1}(\bar{\underset{\sim}{X}}_1-\bar{\underset{\sim}{X}}_2) \quad (3)$$

The observations are classified as belonging to S_1 if $W > 0$; as belonging to S_2 if $W < 0$.

Quadratic Discriminant Function

Table 2 illustrates the means and variances of the three non-normal distributions. Clearly, the variances for S_1 are markedly different from that for S_2. Therefore, it was appropriate to consider classification according to the Quadratic Discriminant Function. The form of the QDF, assuming equal a priori probabilities and equal costs of misclassification is:

$$QDF = 1/2\underset{\sim}{Z}'(\underset{\sim}{S}_1^{-1}-\underset{\sim}{S}_2^{-1})\underset{\sim}{Z} + (\bar{\underset{\sim}{X}}_2\underset{\sim}{S}_2^{-1} - \bar{\underset{\sim}{X}}_1\underset{\sim}{S}_1^{-1})\underset{\sim}{Z} + 1/2([\bar{\underset{\sim}{X}}_1\underset{\sim}{S}_1^{-1}\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2\underset{\sim}{S}_2^{-1}\bar{\underset{\sim}{X}}_2] - \log[\underset{\sim}{S}_2/\underset{\sim}{S}_1]) \quad (4)$$

The new observations are classified into S_1 if $Q > 0$; into S_2 if $Q < 0$.

Nearest Neighbor with Probability Blocks

The Nearest Neighbor with Probability Blocks procedure is based on distribution-free tolerance regions. To obtain the necessary probability blocks, a procedure outlined by Gessaman and Gessaman (1972) was employed. Assume, without loss of generality that $p = 2$--the generalization to general p space is immediate. Let

TABLE 2

MEANS AND VARIANCES OF THE NON-NORMAL DISTRIBUTIONS FOR SPECIFIED MEANS OF THE NORMAL DISTRIBUTION

$(\sigma_y^2 = 1)$[a]

Log Normal		
μ_y	η_x	σ_x^2
0	1.65	4.67
1	4.48	34.51
2	12.18	255.02
3	33.12	1884.32
Logit Normal		
μ_y	η_x	σ_x^2
0	.50	.043
1	.70	.029
2	.84	.019
3	.94	.001
Inverse Hyperbolic Sine Normal		
μ_y	η_x	σ_x^2
0	0	3.19
1	1.94	9.65
2	5.98	74.63
3	16.52	471.94

SOURCE: Lachenbruch, Sneeringer and Revo. Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-Normality. Communications in Statistics, 1973, 1, 54.

[a] σ_y^2 is the variance of the underlying normal distribution.

μ_y is the mean of the underlying normal distribution.

η_x is the mean of the transformed non-normal variate.

σ_x^2 is the variance of the transformed non-normal variate.

$k = k_n = [n^{(p-1)/(p+1)}]$, the greatest integer less than or equal to $n^{(p-1/p+1)}$.

The observations were ranked along the first coordinate and the plane partitioned into $[(n/k)^{1/2}]$ "blocks" by making $[(n/k)^{1/2}]$ -1 evenly spaced "cuts" on the ranked observations. The cut-point belongs to the right boundary of the block of which it forms. Since the distributions are absolutely continuous, ties occur with probability zero.

The observations used to make the cuts were deleted. Then taking each block, the remaining observations were partitioned into $[(n/k)^{1/2}]$ subblocks by making $[(n/k)^{1/2}] - 1$ evenly spaced cuts on the second coordinate. The plane was then partitioned into $[(n/k)^{1/2}]$ probability blocks, each containing k-1, k, or k+1 observations.

Once the probability blocks were determined, the observations in X_1 and X_2 used to develop them were classified into the blocks. A block was considered to be an X_1 block if the majority of the observations in the block were X_1 observations; an X_2 block if the majority of the observations in the block were X_2 observations. If a block had an equal number of observations from both populations, the block was classified according to its neighboring blocks.

Once the membership of each of the blocks was ascertained, the new observations were classified and the number of misclassifications determined.

Density Estimators

For $\hat{f}_1(x)$ and $\hat{f}_2(x)$ consistent estimators of $f_1(x)$ and $f_2(x)$, respectively, the procedure used was the ratio of the density esti-

mators. The new observations are classified into S_1 if the ratio is greater than one; into S_2 if the ratio is less than one.

Parzen-Cacoullos Density Estimator

$$\hat{f}_i(x) = \frac{1}{nh^p(n)} \sum_{j=1}^{n} K\left(\frac{x-y_j}{h(n)}\right) \quad (5)$$

where $h(n) = n^{-1/8}$ and $K(w) = \mathrm{EXP}(-[w_1^2 + \ldots + w_p^2]/2)/(2\pi)^{1/2}$

Loftsgaarden-Quesenberry Density Estimator

$$\hat{f}_i(x) = \frac{k_n - 1}{nA_{r_{k_{n,Z}}}} \quad (6)$$

where $k_n = n^{1/2}$ and $A_{r,Z} = \frac{2r^p\pi^{p/2}}{p\Gamma(p/2)}$ and r is the distance from the new observation to the k_nth closest x_i as determined by Euclidean distance.

Gessaman Density Estimator

$$\hat{f}_i(x) = \frac{k_n}{(n+1)D_{Z,n}} \quad (7)$$

where $k_n = [n^{(p-1)/(p+1)}]$ and D is the area of the bounded block into which the observations falls. When the observation falls into an unbounded block, the Nearest Neighbor procedure is used to classify the observation.

Procedures

Each of the 72 combinations of sample size, dimensionality, mean vector, and distribution were used to form classification procedures for the six discrimination rules. Then 500 new observations from each of the populations were generated and classified according to the rules established. Ten iterations of the process were performed when n = 64 or n = 200 and five iterations were performed

when n = 729.[2] The proportion of misclassified observations from each population and the total proportion of misclassified observations were determined and compared.[3] All computer programs to generate the data and classification procedures were written in the FORTRAN IV programming language.

Once the proportion of misclassification was obtained, there were certain hypotheses tested. For two multivariate normal distributions with identical variance-covariance matrices, the optimal classification procedure is such that the respective probabilities of misclassification are equal (Anderson, 1951). Because of this, the first hypothesis concerned the empirical probabilites of misclassification from each population for each of the six procedures (H_o: $P[1/2] = P[2/1]$).

The second hypothesis concerned the overall probability of misclassification for the procedures. For each of the parameter combinations, the hypothesis of equality of the six proportions was tested. When the hypothesis was rejected, the Marscuilo (1966) analogue to the Scheffe multiple comparison theorem was used to determine significant pairwise contrasts.

A similar testing procedure was used to determine if significant differences existed between the overall proportions of misclassification for the three sample sizes.

Results

Log Normal Distribution (p = 2)

The results for the two dimensional Log Normal distributed

---

[2]Because of the computer time necessary for the iterations when n = 729, and because of a computer cost factor, it was not feasible to perform more than five iterations.

[3]$P(I/J)$ is the proportion of observation from S_J misclassified into S_I.

random variables are presented in Table 3. The Log Normal distribution is an example of a semi-infinite range distribution.[4]

$\underline{\mu = 1}$. The performance of the LDF and QDF was significantly inferior to the four nonparametric procedures. However, there was no significant difference between the overall proportions of misclassification for the four nonparametric procedures except when n = 64. When n = 64, the overall proportion of misclassification for the Nearest Neighbor and Gessaman techniques was significantly smaller than for the Parzen-Cacoullos and Loftsgaarden-Quesenberry procedures. For all procedures there was no significant difference between the overall proportions of misclassification when n = 200 or n = 729. However, there were differences between n = 64 and the other two sample sizes. Therefore, a sample size of 200 was necessary for the criterion sample.

$\underline{\mu = 2}$. The pattern of results when $\mu = 2$ was consistent with the results when $\mu = 1$. The performance of the LDF and the QDF were both significantly worse than the four nonparametric procedures. However, there was no differences between the misclassification proportions for the four nonparametric procedures. A minimally sufficient criterion sample size was n = 200.

$\underline{\mu = 3}$. When n = 200 or n = 729, the LDF's performance was significantly worse than the other five procedures; however, there was no difference between the proportions of misclassification for the other five procedures. When n = 64, the only procedures that were

---

[4]In the tables the following abbreviations were used: LDF--Linear Discriminant Function; QDF--Quadratic Discriminant Function; NN--Nearest Neighbor with Probability Blocks: L-Q--Loftsgaarden-Quesenberry Density Estimator; P-C--Parzen-Cacoullos Density Estimator;GESS--Gessaman Density Estimator.

TABLE 3

RESULTS FOR THE LOG NORMAL DISTRIBUTED RANDOM VARIABLES (p = 2)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.5254	.5480	.5724	.1580	.1292	.1248	.3417	.3386	.3486
	QDF	.5734	.6724	.6996	.1848	.0784	.0836	.3791	.3754	.3916
	NN	.2936	.3514	.3384	.3426	.2912	.2960	.3181	.3213	.3172
	P-C	.4508	.4085	.3784	.2501	.2520	.2758	.3505	.3303	.3271
	L-Q	.3722	.3654	.3508	.3160	.2824	.2888	.3441	.3239	.3198
	GESS	.2882	.3320	.3480	.3312	.3150	.3168	.3097	.3235	.3324
2	LDF	.4710	.4912	.4820	.0626	.0276	.0224	.2668	.2594	.2522
	QDF	.3910	.4038	.4216	.0634	.0408	.0380	.2272	.2223	.2298
	NN	.1832	.1722	.1628	.1676	.1506	.1664	.1754	.1614	.1646
	P-C	.2238	.2046	.1780	.1669	.1388	.1584	.1954	.1717	.1682
	L-Q	.1940	.1944	.1708	.1574	.1344	.1552	.1757	.1644	.1630
	GESS	.1832	.1688	.1752	.1576	.1540	.1572	.1754	.1614	.1662
3	LDF	.4292	.4300	.4488	.0224	.0020	.0012	.2258	.2160	.2250
	QDF	.1618	.1136	.1520	.0448	.0428	.0184	.1053	.0782	.0852
	NN	.0868	.0516	.0540	.1612	.1126	.0864	.1240	.0821	.0702
	P-C	.0938	.0918	.0866	.0621	.0700	.0580	.0780	.0809	.0718
	L-Q	.0896	.0862	.0820	.0514	.0618	.0556	.0705	.0740	.0688
	GESS	.0868	.0516	.0540	.1512	.1126	.0864	.1240	.0821	.0702

equivalent were the Nearest Neighbor and Gessaman and the Parzen-Cacoullos and Loftsgaarden-Quesenberry density estimators. Again, a sample of size n = 200 was sufficiently large.

Logit Normal Distribution (p = 2)

The results for the two dimensional Logit Normal distribution are presented in Table 4. The Logit Normal distribution is a member of the family of finite range distributions.

$\mu = 1$. For the two larger sample sizes, there was no significant difference between any of the six procedures. However, only for the Gessaman procedure when n = 200 and the Nearest Neighbor procedure when n= 729 was the hypothesis of equality of the respective proportions of misclassification (P[1/2] and P[2/1]) not rejected. Therefore, for those situations, the Gessaman and Nearest Neighbor procedures were the most desirable. When n = 64, only the LDF and the Loftsgaarden-Quesenberry procedure were significantly different from each other. A sample size of 200 was sufficiently large for the criterion sample.

$\mu = 2$. The results when $\mu = 2$ resembled those when $\mu = 1$. There was no difference among the six overall proportions of misclassification when n = 64 or n = 729. When n = 200, there was a significant difference; however, there were no significant pairwise contrasts. The Nearest Neighbor and the Gessaman procedures were the most desirous to use because the hypothesis of equality of P(1/2) and P(2/1) was not rejected when n = 200 and n = 729 for those procedures. Again, a sample size of 200 was sufficient.

$\mu = 3$. The results were markedly different when $\mu = 3$. When n = 64 and n = 200 the Loftsgaarden-Quesenberry procedure and the QDF were

TABLE 4

RESULTS FOR THE LOGIT NORMAL DISTRIBUTED RANDOM VARIABLES (p = 2)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.2812	.2820	.2748	.3412	.3482	.3336	.3112	.3151	.3042
	QDF	.2648	.2552	.2476	.3650	.3826	.3672	.3149	.3189	.3074
	NN	.3574	.3068	.3224	.3042	.3464	.2972	.3308	.3266	.3098
	P-C	.2416	.2452	.2332	.3918	.3944	.3848	.3167	.3198	.3090
	L-Q	.3468	.3132	.3228	.3318	.3432	.2908	.3393	.3282	.3068
	GESS	.3488	.3220	.3292	.2932	.3360	.3020	.3210	.3290	.3156
2	LDF	.1074	.1058	.1008	.2442	.2308	.2404	.1758	.1683	.1706
	QDF	.1112	.1138	.1104	.2308	.2174	.2288	.1710	.1656	.1696
	NN	.2002	.1670	.1612	.1490	.1534	.1648	.1746	.1602	.1630
	P-C	.0770	.0778	.0748	.2950	.2800	.2808	.1860	.1789	.1778
	L-Q	.1448	.1456	.1480	.1988	.1778	.1784	.1718	.1617	.1632
	GESS	.2002	.1646	.1588	.1490	.1530	.1696	.1746	.1588	.1642
3	LDF	.0208	.0164	.0204	.1810	.1838	.1916	.1009	.1001	.1060
	QDF	.0552	.0438	.0502	.0866	.0972	.0956	.0709	.0705	.0729
	NN	.1836	.0970	.0876	.0728	.0744	.0580	.1282	.0857	.0728
	P-C	.0122	.0118	.0120	.2308	.2346	.2460	.1215	.1232	.1290
	L-Q	.0404	.0424	.0568	.1236	.1000	.0864	.0820	.0712	.0640
	GESS	.1836	.0970	.0876	.0728	.0744	.0580	.1282	.0857	.0728

the two best performing procedures; when n = 729 all but the LDF and Parzen-Cacoullos procedures were desirable. For most of the procedures a sample of size 200 was sufficient; in some instances a sample of size 64 was large enough.

Inverse Hyperbolic Sine Normal Distribution (p = 2)

The results for the Monte Carlo simulation for the Inverse Hyperbolic Sine Normal distribution when p = 2 are presented in Table 5. The Inverse Hyperbolic Sine Normal distribution is a member of the infinite range family of distributions.

$\mu = 1.$ Except for the QDF, the remaining five procedures were equally effective in classifying the observations for all the samples. Except for the Parzen-Cacoullos density estimator and the Loftsgaarden-Quesenberry density estimator, a sample of size 64 was sufficiently large. For the two density estimators, a sample of size n = 200 was necessary.

$\mu = 2.$ The four nonparametric procedures uniformly misclassified fewer observations than the LDF and QDF for all sample sizes. Since the hypothesis of equality of P(2/1) and P(1/2) was not rejected for the Gessaman procedure (n = 64 or 200) and the Nearest Neighbor procedure (n = 64), in those instances, those procedures were the best procedures to use. Except for the LDF a sample of size n = 200 was sufficient.

$\mu = 3.$ For the two larger sample sizes, the four nonparametric procedures were significantly better than the parametric ones, but not significantly different from each other. The Nearest Neighbor and Gessaman procedures were the best to use in this situation because when n = 200 and n = 729, the hypothesis of equality of P(2/1) and

TABLE 5

RESULTS FOR THE INVERSE HYPERBOLIC SINE NORMAL DISTRIBUTED RANDOM VARIABLES ($p = 2$)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.4260	.4430	.4496	.2192	.2118	.1964	.3226	.3274	.3230
	QDF	.5750	.6228	.6304	.1626	.1232	.1276	.3688	.3730	.3790
	NN	.3474	.3018	.3112	.3156	.3428	.3220	.3315	.3223	.3166
	P-C	.4047	.3626	.3296	.2841	.2958	.3068	.3444	.3292	.3182
	L-Q	.3736	.3510	.3144	.3142	.3026	.3204	.3439	.3268	.3174
	GESS	.3300	.3066	.3118	.3200	.3474	.3240	.3250	.3270	.3214
2	LDF	.4310	.4112	.4664	.0474	.0364	.0332	.2392	.2238	.2498
	QDF	.4286	.4210	.4692	.0358	.0560	.0524	.2572	.2385	.2608
	NN	.1750	.1518	.1880	.1348	.1784	.1416	.1799	.1651	.1648
	P-C	.2428	.2071	.2100	.1312	.1282	.1236	.1870	.1677	.1668
	L-Q	.2370	.2092	.2104	.1298	.1276	.1288	.1834	.1684	.1696
	GESS	.1750	.1584	.1776	.1348	.1642	.1544	.1799	.1613	.1660
3	LDF	.4112	.4056	.4376	.0124	.0054	0024	.2118	.2055	.2200
	QDF	.1776	.1800	.1912	.0456	.0444	.0432	.1116	.1122	.1172
	NN	.1056	.0906	.0656	.1186	.0874	.0668	.1121	.0890	.0662
	P-C	.1108	.1028	.0898	.0645	.0577	.0590	.0877	.0803	.0744
	L-Q	.0978	.0992	.0896	.0572	.0404	.0556	.0775	.0743	.0726
	GESS	.1056	.0906	.0656	.1186	.0874	.0668	.1121	.0890	.0662

P(1/2) was not rejected. The necessary sample size varied with the specific procedure.

Normal Distribution (p = 2)

The results for the two dimensional normally distributed random variables are presented in Table 6. Because all of the assumptions of the LDF are satisfied, it is expected that the LDF would be the optimal procedure in this situation. Hence, the use of the normally distributed random variables serves as a check on the procedures.

$\mu = 1$. As to be expected, the performance of the LDF approached the optimal probability of misclassification. Additionally, as the sample size increased, the performance of the QDF approached the LDF since the variance-covariance matrix approached the identity matrix. There was no difference between the performance of any of the procedures, and a sample of size 64 was sufficiently large to develop an efficient classification rule.

$\mu = 2$. Similar to the results when $\mu = 1$, there were no significant differences among the six procedures. Again a sample size of $n = 64$ was sufficient.

$\mu = 3$. The performance of the procedures was equivalent except for the proportions of the Nearest Neighbor and Gessaman procedures which were significantly worse than the other procedures. Except for the Nearest Neighbor and Gessaman procedures, a sample of size 64 was sufficient; for those two procedures a sample size of 729 was necessary.

Log Normal Distribution (p = 3)

The results for the three dimensional Log Normal random variables are presented in Table 7.

TABLE 6

RESULTS FOR THE NORMALLY DISTRIBUTED RANDOM VARIABLES (p = 2)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.2926	.3130	.3100	.3286	.3018	.3268	.3106	.3074	.3184
	QDF	.2914	.3074	.3056	.3322	.3062	.3312	.3118	.3068	.3184
	NN	.3362	.3522	.3224	.3048	.3086	.3316	.3205	.3304	.3270
	P-C	.2896	.3228	.3148	.3506	.2968	.3260	.3201	.3098	.3204
	L-Q	.3034	.3204	.3244	.3534	.3150	.3476	.3284	.3177	.3360
	GESS	.2958	.3438	.3404	.3442	.3152	.3292	.3200	.3295	.3348
2	LDF	.1710	.1788	.1584	.1600	.1534	.1560	.1655	.1661	.1572
	QDF	.1716	.1772	.1580	.1600	.1540	.1560	.1658	.1656	.1570
	NN	.1638	.1722	.1740	.1854	.1616	.1484	.1746	.1669	.1612
	P-C	.1672	.1808	.1632	.1712	.1514	.1560	.1692	.1661	.1596
	L-Q	.1722	.1810	.1772	.1730	.1544	.1512	.1726	.1677	.1642
	GESS	.1638	.1722	.1704	.1854	.1598	.1652	.1746	.1660	.1678
3	LDF	.0734	.0684	.0652	.0708	.0674	.0640	.0721	.0679	.0646
	QDF	.0718	.0694	.0640	.0716	.0662	.0644	.0717	.0678	.0642
	NN	.1736	.0948	.0672	.0540	.0810	.0756	.1138	.0879	.0714
	P-C	.0794	.0694	.0652	.0706	.0660	.0636	.0750	.0677	.0644
	L-Q	.0778	.0660	.0688	.0716	.0692	.0616	.0747	.0676	.0652
	GESS	.1736	.0948	.0672	.0540	.0810	.0756	.1138	.0879	.0714

TABLE 7

RESULTS FOR THE LOG NORMAL DISTRIBUTED RANDOM VARIABLES (p = 3)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.5390	.5406	.5468	.1664	.1426	.1500	.3527	.3416	.3484
	QDF	.6160	.6660	.6876	.1390	.0944	.0908	.3775	.3802	.3892
	NN	.3500	.2736	.3744	.3504	.3634	.3056	.3502	.3185	.3400
	P-C	.4314	.4043	.3932	.2752	.2816	.2910	.3533	.3430	.3421
	L-Q	.4160	.3688	.3588	.2728	.2780	.2908	.3444	.3234	.3248
	GESS	.3512	.2650	.3624	.3480	.3690	.3176	.3496	.3170	.3400
2	LDF	.4556	.4520	.4665	.0712	.0336	.0245	.2634	.2428	.2465
	QDF	.3866	.3698	.4105	.0814	.0652	.0445	.2340	.2175	.2275
	NN	.2308	.1842	.1590	.1880	.1950	.1585	.2094	.1896	.1585
	P-C	.2473	.2177	.2095	.1520	.1566	.1423	.1997	.1872	.1759
	L-Q	.2274	.2160	.1975	.1328	.1372	.1315	.1801	.1766	.1645
	GESS	.2306	.1842	.1590	.1830	.1950	.1585	.2093	.1896	.1588
3	LDF	.4680	.4514	.4475	.0340	.0134	.0025	.2510	.2324	.2250
	QDF	.1324	.1428	.1450	.0674	.0436	.0130	.0999	.0932	.0790
	NN	.1900	.2062	.1125	.1534	.1220	.0945	.1717	.1641	.1035
	P-C	.1093	.1018	.0965	.0801	.0734	.0780	.0947	.0876	.0823
	L-Q	.1010	.0882	.0945	.0518	.0594	.0610	.0764	.0738	.0778
	GESS	.1890	.2062	.1125	.1528	.1220	.0945	.1709	.1641	.1035

$\mu = 1$. Except for the QDF, the remaining five procedures were equivalent in terms of their overall proportion of misclassified observations. When n = 64, the hypothesis of equality of P(2/1) and P(1/2) was not rejected for the Nearest Neighbor or Gessaman procedures (it was rejected for the other procedures). However, because of the significance between both the Nearest Neighbor and Gessaman procedures based on differing sample sizes, it appeared that even with as large a sample size as n = 729, there was instability in the two procedures. For the LDF, QDF, and Parzen-Cacoullos procedures a sample of size 64 was sufficient; for the Loftsgaarden-Quesenberry procedure a sample of size 200 was necessary.

$\mu = 2$. The performance of the four nonparametric procedures was significantly better than the two parametric procedures. The hypothesis of equality of P(1/2) and P(2/1) was not rejected for n = 200 or n = 729 for the Nearest Neighbor and Gessaman procedures. For those two procedures a sample of at least 729 was necessary. For the remaining procedures a sample of size 200 was sufficient; for the Loftsgaarden-Quesenberry procedure a sample of size 64 was sufficient.

$\mu = 3$. The Loftsgaarden-Quesenberry procedure was uniformly the best procedure to use while the LDF was uniformly the worst. A sample of n = 64 was sufficient for the Loftsgaarden-Quesenberry procedure. For the remaining procedures a sample of size 200 was sufficient; for the Nearest Neighbor and Gessaman procedures, a sample of size 729 was necessary.

Logit Normal Distribution (p = 3)

The three dimensional Logit Normal results are presented in

Table 8.

$\underline{\mu = 1.}$ Unlike the log normal distribution's results, all of the proportions of misclassification in this situation were equivalent. Because for the two larger samples the hypothesis of equality of P(2/1) and P(1/2) was not rejected for the Gessaman procedure, that procedure was the optimal procedure to use. When n = 729, the hypothesis was also not rejected for the Loftsgaarden-Quesenberry procedure; therefore, in that situation, the Loftsgaarden-Quesenberry procedure would be desirous to use.

$\underline{\mu = 2.}$ When n = 64 or n = 200 any of the procedures except for the Nearest Neighbor and Gessaman procedures would be appropriate to use. When n = 729, the Nearest Neighbor and Gessaman procedures were the best to use. For all of the procedures except for the Nearest Neighbor and Gessaman procedures a sample of size 64 was sufficient; for those two procedures a sample of size 729 was necessary.

$\underline{\mu = 3.}$ When n = 64, the QDF was the best procedure to use. However, since the proportion of misclassification when n = 64 was significantly worse than when n = 729, a sample of size n = 64 was not large enough. Similarly, when n = 200 either the QDF or Loftsgaarden-Quesenberry procedures were the best of the six. However, since the procedure when n = 200 was different than the procedure when n = 729 for the Loftsgaarden-Quesenberry procedure, a sample of size 200 was not sufficiently large.

Inverse Hyperbolic Sine Normal Distribution (p = 3)

The results of the three dimensional Inverse Hyperbolic Sine Normal distribution are presented in Table 9.

$\underline{\mu = 1.}$ When n = 64 or n = 200 any of the procedures except for the QDF and Parzen-Cacoullos procedures was appropriate for use. When

TABLE 8

RESULTS FOR THE LOGIT NORMAL DISTRIBUTED RANDOM VARIABLES (p = 3)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.2778	.2860	.2888	.3684	.3360	.3352	.3231	.3110	.3120
	QDF	.2560	.2632	.2632	.3980	.3714	.3664	.3270	.3173	.3148
	NN	.3258	.3106	.3108	.3752	.3374	.3292	.3505	.3240	.3200
	P-C	.2332	.2464	.2528	.4240	.3862	.3760	.3286	.3163	.3144
	L-Q	.3018	.3078	.3120	.3976	.3524	.3280	.3497	.3301	.3200
	GESS	.3284	.3136	.3156	.3690	.3302	.3232	.3487	.3219	.3194
2	LDF	.1152	.1042	.1004	.2304	.2350	.2460	.1728	.1696	.1732
	QDF	.1266	.1138	.1048	.2204	.2242	.2360	.1735	.1690	.1704
	NN	.2184	.1688	.1752	.2074	.2146	.1488	.2129	.1917	.1620
	P-C	.0892	.0792	.0760	.2774	.2734	.2920	.1833	.1763	.1840
	L-Q	.1360	.1384	.1392	.2194	.1964	.1948	.1777	.1674	.1670
	GESS	.2184	.1688	.1664	.2074	.2146	.1512	.2129	.1917	.1588
3	LDF	.0180	.0186	.0152	.1836	.1790	.1516	.1008	.0988	.0834
	QDF	.0568	.0458	.0436	.0938	.0930	.0768	.0753	.0694	.0602
	NN	.1666	.1466	.0124	.1792	.1672	.1688	.1729	.1569	.0906
	P-C	.0096	.0122	.0084	.2506	.2222	.1912	.1301	.1172	.0998
	L-Q	.0256	.0306	.0320	.1708	.1320	.0880	.0982	.0813	.0600
	GESS	.1666	.1466	.0124	.1788	.1672	.1688	.1727	.1569	.0906

TABLE 9

RESULTS FOR THE INVERSE HYPERBOLIC SINE NORMAL DISTRIBUTED RANDOM VARIABLES (p = 3)

Mean	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
1	LDF	.3890	.4338	.4392	.2838	.2240	.2008	.3364	.3289	.3200
	QDF	.4862	.5786	.5736	.2524	.1668	.1468	.3693	.3727	.3602
	NN	.3614	.3328	.3156	.3432	.3188	.3212	.3523	.3258	.3184
	P-C	.3772	.3848	.3562	.3572	.3200	.3204	.3672	.3524	.3383
	L-Q	.3540	.3648	.3424	.3570	.3000	.3084	.3605	.3324	.3254
	GESS	.3600	.3332	.3220	.3430	.3156	.3208	.3515	.3244	.3214
2	LDF	.4190	.4232	.4188	.0764	.0424	.0348	.2477	.2328	.2268
	QDF	.3872	.4302	.4428	.0932	.0742	.0600	.2402	.2522	.2514
	NN	.2036	.1288	.1624	.2224	.2604	.1608	.2130	.1946	.1616
	P-C	.2242	.2277	.2172	.1770	.1515	.1536	.2006	.1896	.1854
	L-Q	.2360	.2278	.2336	.1380	.1262	.1116	.1870	.1770	.1726
	GESS	.2034	.1288	.1640	.2224	.2604	.1592	.2129	.1946	.1616
3	LDF	.3828	.3908	.4236	.0242	.0062	.0020	.2035	.1985	.2128
	QDF	.1584	.1752	.1784	.0616	.0478	.0376	.1100	.1115	.1080
	NN	.2544	.2422	.1064	.1046	.0794	.0892	.1795	.1608	.0978
	P-C	.0980	.1034	.0698	.0752	.0751	.0766	.0866	.0893	.0732
	L-Q	.1086	.1018	.0740	.0484	.0484	.0632	.0785	.0751	.0686
	GESS	.2540	.2424	.1060	.1032	.0782	.0892	.1786	.1603	.0976

n = 729 any of the procedures except for the QDF was adequate. For the two larger sample sizes, the Nearest Neighbor and Gessaman procedures were the best procedures to use because the hypothesis of equality of the respective proportions of misclassification was not rejected. For the four nonparametric procedures a sample of size 200 was necessary; for the QDF and LDF a sample of size 64 was sufficient.

$\mu = 2$. When n = 64 the Loftsgaarden-Quesenberry procedure was the optimum to use; when n = 200 either the Loftsgaarden-Quesenberry or the Parzen-Cacoullos; when n = 729 the Nearest Neighbor and the Gessaman procedure. The parametric procedures were significantly worse than the nonparametric ones.

$\mu = 3$. The Loftsgaarden-Quesenberry and Parzen-Cacoullos procedures were the best procedures to use in this situation. For this situation, a sample of at least 729 was necessary.

Normal Distribution (p = 3)

The results of the three dimensional normally distributed random variables appear in Table 10.

$\mu = 1$. Because the assumptions of the LDF were satisfied, it was the best procedure to use in this situation. However, similar to the case when p = 2, any of the procedures could be effectively used. A sample of size n = 64 was sufficient.

$\mu = 2$. Again, because the assumptions of the LDF were satisfied, the LDF was the optimal procedure to use. When n = 64 or n = 200, any of the procedures except for the Nearest Neighbor and Gessaman procedures would be effectively substituted for the LDF. When

TABLE 10

RESULTS FOR THE NORMALLY DISTRIBUTED RANDOM VARIABLES (p = 3)

...n	Procedure	$\hat{P}(2/1)$			$\hat{P}(1/2)$			$\frac{\hat{P}(1/2) + \hat{P}(2/1)}{2}$		
	N	64	200	729	64	200	729	64	200	729
	LDF	.3106	.3070	.3140	.3238	.3026	.3036	.3172	.3048	.3088
	QDF	.3386	.3090	.3080	.3256	.3128	.3148	.3321	.3109	.3114
	NN	.3084	.3042	.3092	.3746	.3380	.3388	.3415	.3211	.3240
	P-C	.3442	.3168	.3008	.3338	.3336	.3348	.3390	.3252	.3178
	L-Q	.3392	.3190	.3068	.3268	.3228	.3344	.3330	.3209	.3206
	GESS	.3154	.3120	.3164	.3552	.3328	.3344	.3403	.3224	.3254
	LDF	.1650	.1710	.1612	.1576	.1666	.1508	.1613	.1688	.1560
	QDF	.1676	.1704	.1612	.1620	.1672	.1500	.1648	.1688	.1556
	NN	.1746	.2504	.2032	.2466	.1398	.1092	.2106	.1951	.1562
	P-C	.1772	.1680	.1616	.1716	.1782	.1588	.1744	.1731	.1602
	L-Q	.1744	.1682	.1584	.1674	.1736	.1528	.1709	.1709	.1556
	GESS	.1742	.2504	.2008	.2456	.1398	.1092	.2099	.1951	.1550
	LDF	.0632	.0684	.0676	.0710	.0716	.0624	.0671	.0700	.0650
	QDF	.0628	.0670	.0664	.0734	.0700	.0608	.0681	.0685	.0636
	NN	.1848	.1910	.1176	.1906	.1412	.0768	.1877	.1661	.0972
	P-C	.0690	.0728	.0696	.0790	.0704	.0656	.0740	.0716	.0676
	L-Q	.0732	.0664	.0660	.0696	.0738	.0656	.0514	.0701	.0658
	GESS	.1840	.1910	.1176	.1902	.1408	.0768	.1871	.1659	.0972

$\underline{\mu = 3.}$ There was no difference between the performance of any of the six procedures. A sample of size n = 64 was sufficient to obtain consistent results with the optimal proportion of misclassification.

Conclusions

This study has shown that when observations are drawn from non-normal distributions, certain nonparametric discrimination procedures more appropriately classify the observations than does the parametric LDF (or QDF). Even when the data to be classified is from multivariate normal distributions with equal variance-covariance matrices, the performance of certain of the nonparametric procedures parallels that of the parametric procedures. Therefore, usage of the nonparametric procedures is mandated regardless of the distribution functions describing the data.

The following are comprehensive summaries and conclusions drawn from the results of the study concerning the six types of classification procedures under consideration.

Linear Discriminant Function

Because of the theoretical development of the Linear Discriminant Function, the performance of the LDF was most superior for the data from the multivariate normal distributions with equal variance-covariance matrices. For multivariate normally distributed random observations, a sample size of 64 was sufficient to effect an appropriate classification rule for the LDF.

For non-normal distributed random data, the use of the LDF was not appropriate. The LDF's overall proportion of misclassification was largest of the six types of classification procedures for the Log Normal and Inverse Hyperbolic Sine Normal distribution. For

the Logit Normal distribution, the performance of the LDF was comparable to the nonparametric procedures. Additionally, for all of the three non-normal distributions, there was an extreme inflation-deflation effect concerning the respective proportions of misclassification for each population; one of the proportions was much larger than the optimal level, and one was much lower.

For data whose distribution is unknown, it would be unwise to make use of the Linear Discriminant Function to classify the observations.

Quadratic Discriminant Function

The Quadratic Discriminant Function is the analogue to the LDF when the variance-covariance matrices are unequal. For the results of the study based on the multivariate normally distributed random variables, the observations were drawn from multivariate normal distributions with equal variance-covariance matrices. Since the QDF, given equal variance-covariance matrices becomes the LDF, the performance of the QDF paralleled that of the LDF--especially as the sample size increased for the criterion sample (since the estimated parameters more closely resemble the populaton parameters).

Similar to the LDF, it would be unwise to use the QDF to classify observations from unknown distributions.

Nearest Neighbor Procedure with Probability Squares

When the dimensionality was two, the Nearest Neighbor procedure performed well for all of the distributions. There was no discernable difference between the performance of the Nearest Neighbor procedure for any of the distributions.

When p = 3, the performance of the Nearest Neighbor procedure

declined significantly. This decline was not due to the mode of classification of the procedure, nor to an inherent fault in the procedure, but instead due to the development of the probability blocks and the size of the criterion sample.

The number of probability blocks is a function of the sample size; the specific function established so that there is a sufficient range of observations in each block. When $p = 2$, the block development function was sufficient (there are enough observations in each block so that the range of observations that would belong to a particular block was widespread). Hence, the Nearest Neighbor procedure when $p = 2$ was well developed. However, when $p = 3$, because of the desire to maintain the same three criterion sample sizes that were used when $p = 2$, there was no block development function that would effect as appropriate a set of probability blocks as when $p = 2$. For this reason, the Nearest Neighbor procedure when $p = 3$ was less effective than when $p = 2$.

When $p = 2$, in most cases there was a significant difference between the mean proportion when $n = 64$ and that when $n = 200$ or $n = 729$. For this reason, the use of the procedure for small samples may not be desirable. However, because of the desirable property that the Nearest Neighbor procedure is completely distribution-free, because of the relative ease with which it can be developed, and because of its generally good performance, the use of the Nearest Neighbor procedure is appropriate and suggested for any type of classification problem for which the underlying distribution is unknown.

Parzen-Cacoullos Density Estimator

approximates that of the LDF and QDF since the Parzen-Cacoullos estimator is asymptotically a multivariate normal density (Parzen, 1962). However, because of its nonparametric features, the performance of the Parzen-Cacoullos procedure was somewhat better than the LDF and QDF.

The performance of the Parzen-Cacoullos density estimator was equivalent for each of the three non-normal distributions, and as to be expected, its performance was best for the multivariate normally distributed random variables. For the non-normal distributions, the performance of the Parzen-Cacoullos density estimator procedure was not significantly different from the other nonparametric procedures.

Loftsgaarden-Quesenberry Density Estimator

In addition to the Nearest Neighbor procedure, the Loftsgaarden-Quesenberry procedure was that procedure of the six which most effectively classified the observations. There was no difference between the discriminatory power of the Loftsgaarden-Quesenberry procedure for any of the four distributions.

For the Nearest Neighbor procedure, when $n = 64$, the mean proportion of misclassified observations was in general significantly different than when $n = 200$ or $n = 729$, suggesting that a criterion sample size of 64 was not sufficient for the Nearest Neighbor procedure. However, the respective mean proportions of misclassification for the Loftsgaarden-Quesenberry procedure when $n = 64$ was not different from the mean proportion when $n = 200$ or $n = 729$.

When the criterion sample is small, the use of the Loftsgaarden-Quesenberry estimator is suggested; when larger samples are available

to obtain a criterion, either the Nearest Neighbor or the Loftsgaarden-Quesenberry procedure would suffice.

Gessaman Density Estimator

The Gessaman procedure presupposes the existence of probability blocks. Therefore, because the Nearest Neighbor with probability blocks procedure has been shown to be such an effective classification procedure regardless of the type of distribution from which the observations come, it would appear that the use of the Gessaman procedure is unnecessary.

For the instances in which the distributions under consideration were widely separated, the Gessaman procedure became almost identical to the Nearest Neighbor procedure; for the cases in which the probabilities of misclassification for the Gessaman procedure are different from that for the Nearest Neighbor procedure, at no time is the mean proportion of misclassification for the Gessaman procedure significantly less than the mean proportion of misclassification for the Nearest Neighbor procedure.

Summary

In general, based on the results of this study, the Nearest Neighbor and Loftsgaarden-Quesenberry classification procedures were the two types of procedures which uniformly best classified observations from unknown distributions. These two discrimination techniques should be considered as viable alternatives to the parametric Linear Discriminant Function, especially when the distributions of the observations are unknown.

REFERENCES

Anderson, T. W. Classification by multivariate analysis. Psychometrika, 1951, 16, 31-50.

Anderson, T. W. An introduction to multivariate statistical analysis. New York: John Wiley, 1958.

Cacoullos, T. Estimation of a multivariate density. Annals of the Institute of Statistics and Mathematics, Tokyo, 1966, 18, 179-186.

Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 1936, 7, 179-188.

Fix E. & Hodges, J. L. Nonparametric discrimination: Consistency properties. Lackland Air Force Base, Tex.: U. S. School of Aviation Medicine, 1951, (No. 21-49-004, Report No. 4).

Gessaman, M. P. A consistent nonparametric multivariate density estimator based on statistically equivalent blocks. Annals of Mathematical Statistics, 1970, 41, 1344-1346.

Gessaman, M. P. & Gessaman, P. H. A comparison of some multivariate discrimination procedures. Journal of the American Statistical Association, 1972, 67, 468-472.

Hoel, P. G. & Peterson, R. P. A solution to the problem of optimum classification. Annals of Mathematical Statistics, 1949, 20, 433-438.

Johnson, N. L. Systems of frequency curves generated by methods of translation. Biometrika, 1949, 36, 149-176.

Lachenbruch, P. A., Sneeringer, C. & Revo, T. Robustness of the linear and quadratic discriminant function to certain types of non-normality. Communications in Statistics, 1973, 1, 39-57.

Loftsgaarden, D. O. & Quesenberry, C. P. A nonparametric estimate of a multivariate density function. Annals of Mathematical Statistics, 1965, 36, 1049-1051.

MacLaren, M. D. & Marsaglia, G. Uniform random number generator. Journal of the Association of Computing Machinery, 1965, 12, 83-89

Marascuilo, L. A. Large sample multiple comparisons. Psychological Bulletin, 1966, 65, 280-290.

Parzen, E. On estimation of a probability density function and mode. Annals of Mathematical Statistics, 1962, 33, 1065-1076.

Von Mises, R. On the classification of observation data into distinct groups. Annals of Mathematical Statistics, 1945, 16, 68-73.